REMARKS

Claims 1, 3-6, 10-19, and 28-39 remain in the application for consideration. In view of the following remarks, Applicant respectfully requests reconsideration and allowance of the subject application.

§ 102/103 Rejections

Claims 1, 3-6, 14-19, 28-30, 32, 33, and 35-39 stand rejected under 35 U.S.C. §102(e) as being anticipated by US. Patent No. 6,188,976 to Ramaswamy et al. (hereinafter "Ramaswamy").

Claims 9-13, 31 and 34 stand rejected under 35 U.S.C. §103(a) as being unpatentable under Ramaswamy in view of U.S. Patent No. 6,317,707 to Bangalore et al. (hereinafter "Bangalore").

The Claims

Claim 1 has been amended and, as amended, recites a method of using a tuning set of information to jointly optimize the performance and size of a language model, including (added language appears in the bold italics):

- segmenting at least a subset of a received textual corpus into segments by clustering every N-items of the received corpus into a training unit, wherein resultant training units are separated by gaps, *and wherein N is an empirically derived value based, at least in part, on the size of the received corpus;*
- creating the tuning set from application-specific information;
- (a) training a seed model via the tuning set;
- (b) calculating a similarity within a sequence of the training units on either side of each of the gaps;

- (c) selecting segment boundaries that maximize intra-segment similarity and inter-segment disparity;
- (d) calculating a perplexity value for each segment based on a comparison with the seed model;
- (e) selecting some of the segments based on their respective perplexity values to augment the tuning set;
- iteratively refining the tuning set and the seed model by repeating parts (a) through (e) until a threshold; and
- refining the language model based on the seed model.

In making out a rejection of this claim, the Office argues that Ramaswamy anticipates claim 1. While Applicant disagrees with the Office's rejection and reserves its right to continue to argue that Ramaswamy does not anticipate this claim, this claim has been amended to include the claim language of the previously presented claim 9, which has now been canceled. In light of the current amendment, claim 1 now recites segmenting at least a subset of a received textual corpus into segments by clustering every N-items of the received corpus into a training unit, wherein resultant training units are separated by gaps, *and wherein N is an empirically derived value based, at least in part, on the size of the received corpus*. In light of the current amendments, the Applicant respectfully traverses the Office's rejection.

In the Office's last Office Action, dated 11/16/2005, the Office correctly admitted that Ramaswamy does not teach that "N is an empirically derived value based, at least in part, on the size of the received corpus." Applicant agrees. However, the Office argued that Bangalore teaches this subject matter citing to column 2 lines 59-65 of Bangalore. The Office then reasoned that "it would be obvious to one ordinarily skilled in the art to combine Ramaswamy with Bangalore [with the motivation] to include every item in the clustering process to better improve subsequent clustering results for determining the compactness of a cluster. The applicant strongly disagrees and submits that the Office's has not established a

prima facie case of obviousness. Specifically, the combination of Ramaswamy and Bangalore does not teach or in any way suggest segmenting at least a subset of a received textual corpus into segments by clustering every N-items of the received corpus into a training unit, wherein resultant training units are separated by gaps, ***and wherein N is an empirically derived value based, at least in part, on the size of the received corpus.***

In order to assist the Office in further appreciating this, the excerpt cited by the Office at column 2, lines 59-65 of Bangalore as teaching ***N is an empirically derived value based, at least in part, on the size of the received corpus,*** is reproduced below:

> Based upon the frequencies, ***an N dimensional vector may be built for each input word.*** The number of dimensions N of the frequency vector is a multiple of the total number of context words, the total number of input words and the total number of relations identified by the method 1000. The vector represents grammatical links that exist between the input words and the context words.

This excerpt merely describes the creation of a N dimensional vector for each input word, where the vector represents grammatical links that exist between the input words and the context words. The Office appears to have included this excerpt simply because a variable "N" is derived from the total number of context words, the total number of input words and the total number of relations identified by the method 1000. However, upon even a cursory inspection, the variable "N" from Bangalore is not the same as the variable N as recited above. The variable "N" used in Bangalore is not used for clustering every N-items of the received corpus into a training unit ... ***wherein N is an empirically derived value based, at least in part, on the size of the received corpus.*** As such, the combination of Ramaswamy and Bangalore fails to render obvious the subject matter of claim 1. Accordingly, the Office has failed to make out a *prima facie* case of obviousness. In addition, the Office's motivation for making this combination (i.e. for the better

improvement...) is misplaced insofar as it is too general and could seemingly support any modification of the primary reference. This motivation is lacking in the particularity that is required to make out a *prima facie* case of obviousness. For at least these reasons, this claim is allowable.

Claims 3-6 and 10-19 depend from claim 1 and are allowable as depending from an allowable base claim. These claims are also allowable for their own recited features which, in combination with those recited in claim 1, are neither shown nor suggested by the reference of record either singly or in combination with one another.

Claim 28 has been amended, and as amended recites a modeling agent, including (added language appears in the bold italics):

- a controller, to receive invocation requests to develop a language model from a corpus; and
- a data structure generator, responsive to the controller, to:
  - develop a seed model from a tuning set of information;
  - segment at least a subset of a received corpus, wherein the segments of the received corpus are a clustering of every N items of the received corpus into a training unit, *wherein N is an empirically derived value based, at least in part, on the size of the received corpus,* and the training units are separated by gaps;
  - calculate the similarity within a sequence of training units on either side of each of the gaps;
  - select segment boundaries that improve intra-segment similarity and inter-segment disparity;
  - calculate a perplexity value for each segment;
  - refine the seed model with one or more segments of the received corpus based, at least in part, on the calculated perplexity values;
  - iteratively refine the tuning set with segments ranked by the seed model and in turn iteratively update the seed model via the refined tuning set;

- o filter the received corpus via the seed model to find low-perplexity segments; and
- o train the language model via the low-perplexity segments.

Claim 28 has been amended with the same language that was used to amend claim 1. As such, for the same reasons as discussed with regard to claim 1 above, Ramaswamy and Bangalore, either alone or in combination, do not teach or suggest the subject matter of this claim. Of course, Applicant reserves its right to further argue that Ramaswamy does not anticipate this claim in its pre-amended state. As such, this claim is allowable.

Claims 29-35 depend from claim 28 and are allowable as depending from an allowable base claim. These claims are also allowable for their own recited features which, in combination with those recited in claim 28, are neither shown nor suggested by the reference of record either singly or in combination with one another.

Claim 36 has been amended, and as amended recites a method of jointly optimizing the performance and size of a language model, comprising (added language appears in the bold italics):

- segmenting one or more relatively large language corpora into multiple segments of *N items, wherein N is an empirically derived value based, at least in part, on the size of the received corpus;*
- selecting an initial tuning sample of application-specific data, the initial tuning sample being relatively small in comparison to the one or more relatively large language corpora, wherein the initial tuning sample is used for training a seed model, the seed model to be used for ranking the multiple segments from the language corpora;

- iteratively training the seed model to obtain a mature seed model, wherein the iterative training proceeds until a threshold is reached, each iteration of the training including:
  - o updating the seed model according to the tuning sample;
  - o ranking each of the multiple segments according to a perplexity comparison with the seed model;
  - o selecting some of the multiple segments that possess a low perplexity; and
  - o augmenting the tuning sample with the selected segments;
  - o once the threshold is reached, filtering the language corpora according to the mature seed model to select low-perplexity segments;
  - o combining data from the low-perplexity segments; and
  - o training the language model according to the combined data.

Claim 36 has been amended with the same language that was used to amend claim 1. As such, for the same reasons as discussed with regard to claim 1 above, Ramaswamy and Bangalore, either alone or in combination, do not teach or suggest the subject matter of this claim. Of course, Applicant reserves its right to further argue that Ramaswamy does not anticipate this claim in its pre-amended state. As such, this claim is allowable.

Claims 37-39 depend from claim 36 and are allowable as depending from an allowable base claim. These claims are also allowable for their own recited features which, in combination with those recited in claim 36, are neither shown nor suggested by the reference of record either singly or in combination with one another.

Conclusion

All of the claims are in condition for allowance. Accordingly, Applicant requests a Notice of Allowability be issued forthwith. If the Office's next

anticipated action is to be anything other than issuance of a Notice of Allowability, Applicant respectfully requests a telephone call for the purpose of scheduling an interview.

Respectfully Submitted,

Date: 2/9/06

By: _____
        Lee & Hayes PLLC
        Lance Sadler
        Reg. No. 38,605
        (509) 324-9256